

Studijní program: Informatika a její specializace – INFS  
Specializace: Zpracování přirozeného jazyka (NLP) – N0613A140037S02

Tematické okruhy ke státní závěrečné zkoušce **2024/2025**

**KIV/ADSZ – Analýza dat**

Garant: prof. Dr. Ing. Kolingerová

Prerekvizitní předměty: KMA/MSM, KIV/SU, KIV/VI

---

1. Základní principy statistického popisu vícerozměrných dat, bodové a intervalové odhady základních charakteristik náhodných veličin.
2. Lineární regrese jedné i více proměnných, odvození cenové/pokutové funkce a techniky její minimalizace, odvození gradientní metody, algoritmus gradientního sestupu, problémy a omezení gradientního sestupu; polynomiální regrese; normální rovnice.
3. Logistická regrese, model hypotézy logistické regrese, interpretace výsledků, rozhodovací hranice, klasifikace do více tříd – algoritmus One-vs-All.
4. Support Vector Machines, cíl optimalizace jako alternativní pohled na logistickou regresi, matematický model SVM, hypotéza s bezpečnostním faktorem, jádra.
5. Principy testování statistických hypotéz. Testy o shodě středních hodnot, testy o shodě kovariančních struktur.
6. Snižování dimenze (metoda hlavních komponent - cíle metody, předpoklady metody, odvození a použití metody na příkladech; t-SNE - princip, srovnání s PCA).
7. Shluková analýza, cíle metody, předpoklady metody, odvození a použití metody na příkladech. Optimalizační kritérium K-means, výběr centroidů, volba počtu shluků.
8. Diskriminační analýza, cíle metody, odvození a použití metody na příkladech.
9. Různé možnosti vizualizace kvantitativních dat: scatter ploty, histogramy, boxploty, violin ploty, paralelní souřadnice, ....
10. Vizualizaci hierarchií a relací: node-link diagrams, techniky rozložení uzlů (layout), containment diagrams (např. Treemap), layering (např. icicle plots); maticová vizualizace rozsáhlých grafů.
11. Vizuální manipulace - konkrétní příklady a jejich řešení; volba barevné škály, vizualizace nejistoty.

Studijní program: Informatika a její specializace – INFS  
Specializace: **Zpracování přirozeného jazyka (NLP) – N0613A140037S02**

Tematické okruhy ke státní závěrečné zkoušce **2024/2025**  
**KIV/SDSZ – Signály, digitalizovaný obraz, paralelismus**

Garant: prof. Dr. Ing. Kolingerová

Prerekvizitní předměty: KIV/AZS, KIV/ZVI, KIV/PPR

---

1. Základní v čase spojité signály a jejich vlastnosti, vzorkování, kvantizace, číselné řady a jejich vlastnosti, základní operace.
2. Zpracování signálů v časové oblasti, LTI systémy, popis, impulzní odezva, konvoluce, korelace, autokorelační funkce.
3. Zpracování signálů ve frekvenční oblasti, diskrétní Fourierova transformace, vlastnosti, použití DFT pro zpracování jednorozměrných signálů a digitalizovaných snímků.
4. Paralelismus na úrovni instrukcí, predikce skoků, paměťová závislost, falešné sdílení a transakční paměť (Intel TSX) – jejich princip a význam pro urychlení sekvenčních a konkurenčních částí algoritmů.
5. Srovnání paradigmat prostředků pro multithreading – např. POSIX, C++11/Intel TBB, OpenCL/C++ AMP.
6. Přidělování práce v prostředí s distribuovanou pamětí, možnosti urychlení výpočtu a přiřazení procesů na jednotlivé uzly.
7. Úpravy digitalizovaného obrazu s využitím histogramu, detekce hran, filtrace.
8. Matematická morfologie, dilatace, eroze, operace otevření a uzavření, užití pro odstranění šumu a zjednodušení obrazové scény.
9. Ztenčování a skeletizace.
10. Segmentace obrazu.

Studijní program: Informatika a její specializace – INFS  
Specializace: **Zpracování přirozeného jazyka (NLP)** – N0613A140037S02

Tematické okruhy ke státní závěrečné zkoušce **2024/2025**

**KIV/NLSZ – Zpracování přirozeného jazyka**

Garant: doc. Ing. Pavel Král, Ph.D.

Prerekvizitní předměty: KIV/IR, KKY/NSES a KIV/ANLP

---

1. Modely vyhledávání, mechanismy indexování, metody určení podobnosti dotazu a dokumentu.
2. Klasifikace: metody, výběr příznaků, vyhodnocování úspěšnosti. Aplikace klasifikace ve vyhledávání a zpracování přirozeného jazyka.
3. Shlukování: metody, volba počtu shluků, vyhodnocování úspěšnosti. Aplikace shlukování ve vyhledávání a zpracování přirozeného jazyka.
4. Metody předzpracování textových dat – tokenizace, stemming, lemmatizace.
5. Jazykové modelování – účel a základní metody. Vyhazování modelů.
6. Sémantika slova, distribuční hypotéza, statistické metody zjišťování sémantické podobnosti slov a dokumentů.
7. Neuronové sítě, matematický model umělého neuronu, aktivační funkce, topologie, algoritmus zpětného šíření chyby.
8. Rekurentní neuronové sítě.
9. Konvoluční neuronové sítě.
10. Architektura Transformer, BERT model.